IN THE CLAIMS:
Please revise the claims as follows:

1. (Currently amended) A method of converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data, ~~said representation to be based on a dictionary previously developed for said document corpus and wherein each term in said dictionary has associated therewith a corresponding unique integer,~~ said method comprising:

developing a first vector for said entire document corpus, said first vector being a listing of ~~said unique~~ integers corresponding to ~~dictionary~~ terms in said documents such that each said document in said document corpus is sequentially represented in said listing~~, and~~

~~developing a second vector for said entire document corpus, said second vector indicating the location of each said document's representation in said first vector~~.

2. (Currently amended) The method of claim ~~1, further~~ 19, further comprising:

developing a third vector for said entire document corpus, said third vector comprising a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor.

3. (Original claim) The method of claim 1, further comprising:

rearranging, in said first vector, an order of said unique integers within the data for each said document so that all identical unique integers are adjacent.

4. (Original claim) The method of claim 2, wherein said normalization factor is calculated as:

$$NF = 1/ (\Sigma \ x_i^2)^{1/2}$$, where x_i is the number of occurrences of a specific term in said

document, so that NF represents the reciprocal of the square root of the sum of squares of all

term occurrences in said document.

5. (Currently amended) A method of converting, organizing, and representing in a computer

memory a document corpus containing an ordered plurality of documents, ~~for use by a data~~

~~mining application program requiring occurrence-of-terms data, said representation to be~~

~~based on terms in a dictionary previously developed for said document corpus and wherein~~

~~each said term in said dictionary has associated therewith a corresponding unique integer,~~ said

method comprising:

for said document corpus, taking in sequence each said ordered document and

developing a first uninterrupted listing of ~~said unique~~ integers to correspond to ~~the~~ an

occurrence of ~~said dictionary~~ terms in the document corpus~~; and~~

~~developing a second uninterrupted listing for said entire document corpus, containing~~

~~in sequence the location of each corresponding document in said first uninterrupted listing,~~

~~wherein said first listing and said second listing are provided as input data for said data mining~~

~~application program.~~

6. (Currently amended) The method of claim ~~1, further~~ 21, further comprising:

developing a third uninterrupted listing for said entire document corpus, said third

uninterrupted listing containing a sequential listing of floating point multipliers, each said

floating point multiplier representing a document normalization factor for a corresponding

document in said document corpus.

7. (Original claim) The method of claim 5, further comprising:

for each said document in said document corpus, rearranging said unique integers so

that any identical integers are adjacent.

8. (Original claim) The method of claim 6, wherein said normalization factor is calculated

as:

$NF = 1/ (\Sigma\ x_i^2)^{1/2}$, where x_i is the number of occurrences of a specific term in said

document, so that NF represents the reciprocal of the square root of the sum of squares of all

term occurrences in said document.

9. (Currently amended) An apparatus for organizing and representing in a computer memory

a document corpus containing an ordered plurality of documents, ~~for use by a data mining~~

~~applications program requiring occurrence-of-terms data, said representation to be based on~~

~~terms in a dictionary previously developed for said document corpus and wherein each said~~

~~term in said dictionary has associated therewith a corresponding unique integer,~~ said apparatus

comprising:

an integer ~~determiner~~ <u>determining module</u> receiving in sequence each said

ordered document of said document corpus and developing a first uninterrupted listing of said

unique integers to correspond to ~~the~~ <u>an</u> occurrence of ~~said dictionary~~ terms in the document

corpus; ~~and~~

~~a locator developing a second uninterrupted listing for said entire document corpus containing in sequence the location of each corresponding document in said first uninterrupted listing, wherein said first listing and said second listing are provided as input data for said data mining applications program~~.

10. (Currently amended) The apparatus of claim ~~9~~ 23, further comprising:

a normalizer developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

11. (Original claim) The apparatus of claim 9, further comprising:

a rearranger rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

12. (Original claim) The apparatus of claim 10, wherein said normalizer calculates said normalization factor as:

$NF = 1/ (\Sigma x_i^2)^{1/2}$, where x_i is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

13. (Currently amended) A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a

method to organize and represent in a computer memory a document corpus containing an ordered plurality of documents, ~~for use by a data mining algorithm requiring occurrence-of-terms data, said representation to be based on terms in a dictionary previously developed for said document corpus and wherein each said term in said dictionary has associated therewith a corresponding unique integer,~~ said method comprising:

developing a first uninterrupted listing of said unique integers to correspond to the occurrence of said dictionary terms in the document corpus~~; and~~

~~a second uninterrupted listing for said entire document corpus containing in sequence the location of each corresponding document in said first uninterrupted listing, wherein said first listing and said second listing are provided as input data for said data mining algorithm~~.

14. (Original claim) The signal-bearing medium of claim ~~13, wherein~~ 25, wherein said method further comprises:

developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

15. (Original claim) A data converter for organizing and representing in a computer memory a document corpus containing an ordered plurality of documents, for use by a data mining applications program requiring occurrence-of-terms data, said representation to be based on terms in a dictionary previously developed for said document corpus and wherein each said term in said dictionary has associated therewith a corresponding unique integer, said data converter comprising:

means for developing a first uninterrupted listing of said unique integers to correspond to the occurrence of said dictionary terms in the document corpus and; and

means for developing a second uninterrupted listing for said entire document corpus containing in sequence the location of each corresponding document in said first uninterrupted listing, wherein said first listing and said second listing are provided as input data for said data mining applications program.

16. (Original claim) The data converter of claim 15, further comprising:

means for developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

17. (Original claim) The data converter of claim 15, further comprising:

means for rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

18. (New) The method of claim 1, further comprising:

developing a dictionary comprising said terms contained in said document corpus; and

associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers being said integers comprising said first vector.

19. (New) The method of claim 1, further comprising:

developing a second vector for said entire document corpus, said second vector

indicating the location of each said document's representation in said first vector.

20. (New) The method of claim 5, further comprising:

developing a dictionary comprising terms contained in said document corpus; and

associating, with each said dictionary term, an integer to be uniquely corresponding to

said dictionary term, said uniquely corresponding integers used in said first uninterrupted

listing.

21. (New) The method of claim 5, further comprising:

developing a second uninterrupted listing for said entire document corpus, said second

uninterrupted listing containing, in sequence, the location of each corresponding document in

said first uninterrupted listing.

22. (New) The apparatus of claim 9, further comprising:

a dictionary developing module to develop a dictionary of terms contained in said

document corpus, each said term being associated with a corresponding unique integer.

23. (New) The apparatus of claim 9, further comprising:

a locator module developing a second uninterrupted listing for said entire document

corpus, said second uninterrupted listing containing, in sequence, the location of each

corresponding document in said first uninterrupted listing.

24. (New) The signal-bearing medium of claim 13, wherein said method further comprises:

developing a dictionary comprising terms contained in said document corpus; and

associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers used in said first uninterrupted listing.

25. (New) The signal-bearing medium of claim 13, wherein said method further comprises:

developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.